What is claimed is:

1. A method of automatically identifying a pattern on a page, comprising:

synthetically generating textual patterns as signal templates;

compensating, if necessary, for visual differences between the synthetically generated textual patterns and images being compared against the synthetically generated images; and

comparing compensated images against images in a database.

2. The method of claim 1, comprising outputting a signal against a synthetically generated image.

3. The method of claim 1, wherein said compensating step accommodates for visual differences between font typefaces and different font sizes.

4. The method of claim 1, further comprising deleting a duplicate scanned first page.

5. The method of claim 1, further comprising identifying pages as duplicates and assessing the duplicates for quality and deleting lower quality page of the duplicates.

6. The method of claim 5, comprising performing a connected element analysis to identify speckle and blocks of solid color.

7. The method of claim 1, wherein the said compensating step comprises reducing resolution, inserting and mirroring a page image in the database.

8. The method of claim 7, comprising moving the page image from the spatial domain to a frequency domain.

9. The method of claim 8, comprising reducing resolution, inverting and mirroring image.

10. The method of claim 1, comprising producing a similarity matrix for search pattern locations identified in said comparing step.

11. The method of claim 1, wherein said compensating step can accommodate visual differences between different typefaces, different font sizes and distortions introduced in subsequent printing, handling and/or scanning of the page.

12. The method of claim 1, wherein said compensating step can accommodate visual differences occurring from producing a graphic image.

13. The method of claim 1, comprising creating a database of metadata to use in synthetically generating patterns.

14. The method of claim 1, comprising creating a target to search for using a search word specified using numeric characters in the search word.

15. The method of claim 14, wherein compensations include small enlargements or reductions in search pattern size or visual distortions.

16. A computer software product configured to automatically identify a pattern on a page that includes said computer software product, comprising a medium readable by a processor, the medium having stored thereon:

a first sequence of instructions which, when executed by said processor, causes said processor to:

synthetically generate textual patterns as signal templates;

a second sequence of instructions which when executed by said processor, causes said processor to compensate, if necessary, for visual differences between the synthetically generated textual patterns and images being compared against the synthetically generated images; and

a third set of instructions, which when executed by said processor, causes said processor to compare compensated images against images in a database.

17. An optical apparatus, configured to automatically identify a pattern on a page, comprising:

a generating unit for synthetically generating textual patterns as signal templates;

a compensating unit for compensating, if necessary, for visual differences between the synthetically generated textual patterns and images being compared against the synthetically generated images; and

a comparing unit for comparing compensated images against images in a database.

18. A computer-readable medium configured to automatically identify a pattern on a page, having stored thereon a plurality of sequences of instructions, said plurality of sequences of instructions which, when executed by a processor, cause said processor to perform the steps of:

synthetically generating textual patterns as signal templates;

compensating, if necessary, for visual differences between the synthetically generated textual patterns and images being compared against the synthetically generated images; and

comparing compensated images against images in a database. ·

19. A computer system for automatically identifying a pattern on a page, said computer system comprising a processor and a memory coupled to said processor; the memory having stored therein sequences of instructions, which, when executed by said processor to perform the steps of:

synthetically generating textual patterns as signal templates;

compensating, if necessary, for visual differences between the synthetically generated textual patterns and images being compared against the synthetically generated images; and

comparing compensated images against images in a database.